

# Affiliation networks with an increasing degree sequence

Yong Zhang<sup>\*</sup>Xiaodi Qian<sup>†</sup>Hong Qin<sup>‡</sup>Ting Yan<sup>§</sup>

Central China Normal University

## Abstract

Affiliation network is one kind of two-mode social network with two different sets of nodes (namely, a set of actors and a set of social events) and edges representing the affiliation of the actors with the social events. Although a number of statistical models are proposed to analyze affiliation networks, the asymptotic behaviors of the estimator are still unknown or have not been properly explored. In this paper, we study an affiliation model with the degree sequence as the exclusively natural sufficient statistic in the exponential family distributions. We establish the uniform consistency and asymptotic normality of the maximum likelihood estimator when the numbers of actors and events both go to infinity. Simulation studies and a real data example demonstrate our theoretical results.

**Key words:** Affiliation networks; Asymptotic normality; Consistency; Maximum likelihood estimators

**Mathematics Subject Classification:** 62E20, 62F12.

Running title: Affiliation networks

## 1 Introduction

Affiliation network is one kind of two-mode social network that consists of two different types of sets of nodes, namely, a set of actors and a set of social events. The network edges indicate the affiliation of actors with social events. Such network data are commonly used to represent memberships between social organizations and their members, for example,

---

<sup>\*</sup>Department of Statistics<sup>\*,†,‡,§</sup> and Hubei Key Laboratory of Mathematical Sciences<sup>†,§</sup>, Central China Normal University, Wuhan, 430079, China. **Emails:** <sup>\*</sup>zhang\_yong@mails.ccnu.edu.cn, <sup>†</sup>qianxiaodi@mails.ccnu.edu.cn, <sup>‡</sup>qinhong@mail.ccnu.edu.cn, <sup>§</sup>tingyanty@mail.ccnu.edu.cn.

the affiliation of the researchers to the academic institutions or interlocking directors to companies or actors to movies. Other scenarios include ceremonial events attended by members faculty sit, social events people attend, trade partners of major oil exporting nations and so on.

In affiliation networks, the actors are brought together to jointly participate in social events. Joint participation in events not only provides the opportunity for actors to interact, but also increases the probability that links (e.g., friendship) between actors form. For example, belonging to the same organizations (boards of directors, political party, labor union, and so on) provides the opportunity for people to meet and interact, and thus a link between individuals is more easily to form in these circumstances. Similarly, when actors participate in more than one event, two events are connected through these actors. There has been increasing interest in analyzing affiliated network data in recent years. A number of approaches are proposed to analyze or model affiliation network data [e.g., [Conyon and Muldoon \(2004\)](#), [Robins and Alexander \(2004\)](#), [Snijders et al. \(2013\)](#)]. [Iacobucci and Wasserman \(1990\)](#) proposed a  $p_2$  exponential family distribution using the degree sequence as the sufficient statistics to model the weighted affiliation network, which is a close relative of  $p_1$  model introduced by [Holl and Leinhardt \(1981\)](#). [Latapy et al. \(2008\)](#) extended the basic network statistics used to analyze one-mode networks to give a description of analysis for two-mode networks systematically. [Snijders et al. \(2013\)](#) proposed a stochastic actor-oriented model for the co-evolution of two-mode and one mode networks. By extending exponential random graph models for the one-mode networks, [Wang et al. \(2009\)](#) proposes a number of two-mode specifications as the sufficient statistics in exponential family graph models for two-mode affiliation networks and compared the goodness of fit results obtained using the maximum likelihood and pseudo-likelihood approaches by simulation.

At present, little theoretical results are obtained in affiliation network models although many properties of statistical models for one-mode networks are derived [e.g., [Chatterjee and Diaconis \(2013\)](#), [Shalizi and Rinaldo \(2013\)](#), [Bhattacharyya and Bickel \(2016\)](#)]. Even in the aforementioned simple [Iacobucci and Wasserman \(1990\)](#)'s model by assuming that all edges are independent, the asymptotic properties of the maximum likelihood estimator (MLE) are still not addressed due to a growing dimension of parameter space. In this paper, we study the asymptotic properties of the MLE in an affiliation model with the degree sequence as the exclusively natural sufficient statistic in the exponential family distributions. This model is identical to the [Iacobucci and Wasserman \(1990\)](#) model for unweighed edges (i.e., binary edges). We establish the uniform consistency and asymptotic normality of the maximum likelihood estimator when the number of actors and events both go to infinity. A key step to the proof is that we make use of the approximate inverse of the Fisher information matrix with small approximation errors, which is the extension

of that used in [Yan et al. \(2016a\)](#).

The rest of this paper is organized as follows. We introduce the model in [Section 2](#). In [Section 3](#), we present the asymptotic results including the uniform consistency and asymptotic normality of the MLE. Simulation studies and a real application to the student extracurricular affiliation data are given in [Section 4](#). Some further discussions are devoted to [Section 5](#). All the proofs are putted in the appendix.

## 2 Model

### 2.1 Notations

Let  $A$  be an event set with  $m$  events denoted by  $\{1, \dots, m\}$ , and  $P$  be an actor set with  $n$  actors denoted by  $\{1, \dots, n\}$ . An affiliation network  $\mathcal{G}(m, n)$  records the affiliation of each actor with each event in an affiliation matrix  $X = (x_{i,j})_{m \times n}$ ;  $x_{i,j} = 1$  if actor  $i$  is affiliated with event  $j$  and  $x_{i,j} = 0$  otherwise. Each row of  $X$  describes an actor's affiliation with the events and each column describes the memberships of the event. In practice,  $n$  is usually large and  $m$  relatively small. Therefore, we assume  $m \leq n$  hereafter. The affiliation network  $\mathcal{G}(m, n)$  can also be represented in a bipartite graph, in which the nodes are partitioned into two subsets for the actors and the events and the edges exist only between pairs of nodes belonging to different subsets. In bipartite graphs, no two actors are adjacent and no two events are adjacent. If pairs of actors are reachable, it is only via paths containing one or more events. Similarly, if pairs of events are reachable, it is only via paths containing one or more actors.

Define  $d_i = \sum_{j=1}^n x_{i,j}$  as the degree of vertex  $i \in A$  and  $\mathbf{d} = (d_1, \dots, d_m)^\top$ . Similarly, define  $b_j = \sum_{i=1}^m x_{i,j}$  as the degree of vertex  $j \in P$  and  $\mathbf{b} = (b_1, \dots, b_n)^\top$ . The pair  $\{\mathbf{d}, \mathbf{b}\}$  is the degree sequence of the affiliation network  $\mathcal{G}(m, n)$ .

### 2.2 An affiliation network model

In this subsection, we present an exponential random bipartite graph model for affiliation networks with the degree sequence as the exclusively natural sufficient statistic. The probability mass function on the affiliation network  $\mathcal{G}(m, n)$  is assumed to be of exponential form:

$$P(\mathcal{G}(m, n)) = \exp(\boldsymbol{\alpha}^\top \mathbf{d} + \boldsymbol{\beta}^\top \mathbf{b} - Z(\boldsymbol{\alpha}, \boldsymbol{\beta})), \quad (2.1)$$

where  $Z(\boldsymbol{\alpha}, \boldsymbol{\beta})$  is the normalizing constant,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)^\top$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^\top$  are parameter vectors. Each affiliation network with the same degree sequence is equally judged. The parameter  $\alpha_i$  quantifies the popularity of the event  $i$  and  $\beta_j$  quantifies the

activity of the actor  $j$  to participate in events. Note that

$$\exp(\boldsymbol{\alpha}^\top \mathbf{d} + \boldsymbol{\beta}^\top \mathbf{b}) = \exp\left(\sum_{i=1}^m \sum_{j=1}^n (\alpha_i + \beta_j) x_{i,j}\right) = \prod_{i=1}^m \prod_{j=1}^n \exp((\alpha_i + \beta_j) x_{i,j}),$$

which implies that the  $mn$  random variables  $x_{i,j}$  are mutually independent and  $Z(\boldsymbol{\alpha}, \boldsymbol{\beta})$  can be expressed as

$$Z(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^m \sum_{j=1}^n \log(1 + \exp(\alpha_i + \beta_j)).$$

Therefore,  $x_{i,j}$ ,  $1 \leq i \leq m, 1 \leq j \leq n$  are mutually independent Bernoulli random variables with the success probability:

$$\mathbb{P}(x_{ij} = 1) = \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad (2.2)$$

which is the [Iacobucci and Wasserman \(1990\)](#) model for affiliated networks with binary edges.

Since the sample is just one realization of the bipartite random graph, the density or probability mass function (2.1) is also the likelihood function. If one transforms  $(\boldsymbol{\alpha}, \boldsymbol{\beta})$  to  $(\boldsymbol{\alpha} - c, \boldsymbol{\beta} + c)$ , the likelihood does not change. Following [Yan et al. \(2016a\)](#), we set  $\beta_n = 0$  for the identifiability of the parameter.

The consistency and asymptotic normality of the MLE in the exponential random graph model with the degree sequence for undirected one-model networks or the bi-degree sequence for directed one-mode networks have been established recently [[Chatterjee et al. \(2011\)](#), [Hillar and Wibisono \(2013\)](#), [Yan and Xu \(2013\)](#), [Yan et al. \(2016a\)](#), [Yan et al. \(2016b\)](#)]. For the bipartite networks, the asymptotic theory for the MLE in the model (2.1) has not been explored. The model is closely related to the Rasch model [[Rasch \(1960\)](#)] for dichotomous item response experiments, which assumes that item  $i$  correctly gives a response to subject  $j$  with probability  $\exp(\alpha_i - \beta_j) / (1 + \exp(\alpha_i - \beta_j))$ . By assuming that all the parameters in the Rasch model are bounded, [Haberman \(1977\)](#) proved the consistency and asymptotic normality of the MLE when the number of items and subjects goes to infinity simultaneously.

### 3 Asymptotic results

Let  $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_{n-1})^\top$  and  $\mathbf{g} = (d_1, \dots, d_m, b_1, \dots, b_{n-1})^\top$ . The log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^m \alpha_i d_i + \sum_{j=1}^{n-1} \beta_j b_j - \sum_{i=1}^m \sum_{j=1}^n \log(1 + e^{\alpha_i + \beta_j}).$$

The likelihood equations are:

$$\begin{aligned} d_i &= \sum_{j=1}^n \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad i = 1, \dots, m, \\ b_j &= \sum_{i=1}^m \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad j = 1, \dots, n-1. \end{aligned} \quad (3.1)$$

Let  $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_m, \widehat{\beta}_1, \dots, \widehat{\beta}_{n-1})^\top$  be the MLE of  $\boldsymbol{\theta}$  and  $\widehat{\beta}_n = 0$ . If  $\widehat{\boldsymbol{\theta}}$  exists, then it is the solution to the system of equation (3.1).

Let  $V = (v_{i,j})$  be the Fisher information matrix of the parameter vector  $\boldsymbol{\theta}$ , which is a diagonal dominant matrix with nonnegative entries. The diagonal elements of  $V$  are

$$v_{i,i} = \sum_{j=1}^n \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}, \quad i = 1, \dots, m, \quad v_{m+j, m+j} = \sum_{i=1}^m \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}, \quad j = 1, \dots, n.$$

Motivated by the approximate inverse proposed by Yan et al. (2016a) for the Fisher information matrix in the directed one-mode network model involved with the bi-degree sequence, we proposed a generalized simple matrix  $S = (s_{i,j})$  to approximate the  $V^{-1}$ , which is defined as

$$s_{i,j} = \begin{cases} \frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n, m+n}}, & i, j = 1, \dots, m, \\ -\frac{1}{v_{m+n, m+n}}, & i = 1, \dots, m, \quad j = m+1, \dots, m+n-1, \\ -\frac{1}{v_{m+n, m+n}}, & i = m+1, \dots, m+n-1, \quad j = 1, \dots, m, \\ \frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n, m+n}}, & i, j = m+1, \dots, m+n-1, \end{cases} \quad (3.2)$$

where  $\delta_{i,j} = 1$  when  $i = j$  and 0 otherwise. For a vector  $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ , denote the  $\ell_\infty$  norm of  $\mathbf{x}$  by  $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$ . We present the consistency of  $\widehat{\boldsymbol{\theta}}$  here, whose proof is given in Appendix A.

**Theorem 1.** Assume that  $\boldsymbol{\theta}^* \in \mathbb{R}^{m+n-1}$  with  $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ , where  $\tau \in (0, 1/24)$  is a constant, and that  $X \sim \mathbb{P}_{\boldsymbol{\theta}^*}$ , where  $\mathbb{P}_{\boldsymbol{\theta}^*}$  denotes the probability distribution (2.1) on  $X$  under the parameter  $\boldsymbol{\theta}^*$ . If  $m/n = O(1)$ , then as  $n$  goes to infinity, with probability

approaching one, the MLE  $\hat{\boldsymbol{\theta}}$  exists and satisfies

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{\infty} = O_p \left( \frac{(\log n)^{1/2} e^{6\|\boldsymbol{\theta}^*\|_{\infty}}}{n^{1/2}} \right) = o_p(1).$$

Further, if the MLE exists, it is unique.

Next, we present the central limit theorem of  $\hat{\boldsymbol{\theta}}$ , whose proof is given in Appendix B.

**Theorem 2.** Assume that  $X \sim \mathbb{P}_{\boldsymbol{\theta}^*}$ . If  $m/n = O(1)$  and  $\|\boldsymbol{\theta}^*\|_{\infty} \leq \tau \log n$ , where  $\tau \in (0, 1/36)$  then for any fixed  $k \geq 1$ , as  $n \rightarrow \infty$ , the vector consisting of the first  $k$  elements of  $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$  is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left  $k \times k$  block of  $S^*$ . where  $S^*$  is the matrix by replacing  $\boldsymbol{\theta}$  in  $S$  given in (3.2) with its true value  $\boldsymbol{\theta}^*$ .

**Remark 1.** By Theorem 2, for any fixed  $i$ , as  $n \rightarrow \infty$ , the convergence rate of  $\hat{\theta}_i$  is  $1/v_{i,i}^{1/2}$ . Since  $me^{-2\|\boldsymbol{\theta}^*\|_{\infty}} \leq v_{i,i} \leq n/4$ , the rate of convergence is between  $O(m^{-1/2}e^{\|\boldsymbol{\theta}^*\|_{\infty}})$  and  $O(n^{-1/2})$ .

## 4 Simulation studies

We carry out the numerical simulations to evaluate Theorem 2. Following Yan et al. (2016a), the parameter values take a linear form. Specifically, we set  $\alpha_{i+1}^* = (m-1-i)L/(m-1)$  for  $i = 0, 1, \dots, m-1$ , and considered four different values 0,  $\log(\log(m))$ ,  $\log(m)^{1/2}$  and  $\log(m)$  for  $L$ . For the parameter vector  $\boldsymbol{\beta}$ , let  $\beta_{j+1}^* = (n-1-i)L/(n-1)$ ,  $j = 0, 1, \dots, n-2$  for simplicity and  $\beta_n^* = 0$  by default.

By Theorem 2,  $\hat{\xi}_{i,j} = [\hat{\alpha}_i - \hat{\alpha}_j - (\alpha_i^* - \alpha_j^*)]/(1/\hat{v}_{i,i} + 1/\hat{v}_{j,j})^{1/2}$ ,  $\hat{\eta}_{i,j} = [\hat{\beta}_i - \hat{\beta}_j - (\beta_i^* - \beta_j^*)]/(1/\hat{v}_{m+i,m+i} + 1/\hat{v}_{m+j,m+j})^{1/2}$ , are asymptotically distributed as standard normal distribution, where  $\hat{v}_{i,i}$  is the estimate of  $v_{i,i}$  by replacing  $\boldsymbol{\theta}$  with  $\hat{\boldsymbol{\theta}}$ . We assess the asymptotic normality of  $\hat{\xi}_{i,j}$ ,  $\hat{\eta}_{i,j}$  using the quantile-quantile (QQ) plot. Further, we record the coverage probability of the 95% confidence interval, the length of the confidence interval, and the frequency that the MLE does not exist. Each simulation is repeated 10,000 times.

We only simulate a single combination for  $(m, n)$  with  $m = 100, n = 200$ , and present the QQ-plots of  $\hat{v}_{ii}^{-1/2}(\hat{\alpha}_i - \alpha_i)$  and  $\hat{v}_{ii}^{-1/2}(\hat{\beta}_i - \beta_i)$  in Figure 1 and Figure 2, respectively. The horizontal and vertical axes are the theoretical and empirical quantiles, respectively, and the straight lines correspond to the reference line  $y = x$ . In Figure 1, we can see that the empirical quantiles coincide with the theoretical ones very well. In Figure 2, there are slight deviations when  $L = (\log m)^{1/2}$ . When  $L = (\log n)^{1/2}$ , there are a little derivations on both tails of plots. When  $L = \log n$ , the MLE does not exist in all repetitions, the QQ plots are not available in this case.

The coverage probability of the 95% confidence interval for  $\alpha_i - \alpha_j$  and  $\beta_i - \beta_j$ , the length of the confidence interval, and the frequency that the MLE did not exist, which are reported in Table 1 and Table 2, respectively. There are similar results for the two tables. We can see that the length of estimated confidence interval increases as  $L$  increases for fixed  $n$ , and decreases as  $n$  increases for fixed  $L$ . The coverage frequencies are all close to the nominal level 95%. When  $L = (\log n)^{1/2}$  (conditions in Theorem 2 no longer hold), the MLE does not exist with a positive probability; when  $L = \log(n)$ , the MLE did not exist with 100% frequencies.

*A data example.* We analyze a student extracurricular affiliation network data collected by Dan McFarland in 1996, which can be downloaded from [http://dl.dropbox.com/u/25710348/snaimages/mag\\_act96.txt](http://dl.dropbox.com/u/25710348/snaimages/mag_act96.txt). It consists of 1295 students (anonymized) and 91 student organizations in which they are members (e.g. National Honor Society, wrestling team, cheerleading squad, etc.). In order to guarantee the existence of the MLE, we remove those 438 individuals that don't belong to any organizations. The MLEs of the parameters for remaining students and organizations and their standard errors as well as the 95% confidence intervals are reported in Tables 3 and 4. The value of estimated parameters reflect the size of degrees. For example, the largest five degrees in student organizations are 199, 157, 124, 93, 89 for organizations "Spanish.Club, Pep.Club, NHS, Latin.Club, Orchestra.Symphonic", which also have the top five influence parameters  $-0.32, -0.64, -0.94, -1.92, -1.34$ . On the other hand, the organizations with the five smallest influence parameters  $-4.60, -4.60, -4.60, -4.89, -4.89$  have degrees 4, 4, 4, 3, 3.

## 5 Summary and discussion

Statistical models for affiliation networks provide insight into the formulation of complex social affiliation between actors and events. They also indirectly reflect how events create ties among actors and the actors create ties among events. Meanwhile, the asymptotic inference in these models are challenge like other network models due to that the structure of the network data is non-standard. In this paper, we derive the uniform consistency and asymptotic normality of the MLE in the exponential random bipartite graph models for affiliation networks with the degree sequence as the exclusively sufficient statistic. The conditions imposed on  $\tau$  that guarantee the good asymptotic properties of the MLE may not be best possible. In particular, the conditions guaranteeing the asymptotic normality are stronger than those guaranteeing the consistency. Simulation studies suggest that the conditions on  $\tau$  might be relaxed. The asymptotic behavior of the MLE depends not only on  $\tau$ , but also on the configuration of all the parameters. We will investigate this problem in the future work.

We only consider dyadic independence assumption. Like the model specifications in

the exponential random graph models for the one-mode network, one can add the counts of  $k$ -stars of various sizes as the model terms in the model (2.1) to model the dependence structure. However, such model incurs the degeneracy problem [e.g., Chatterjee and Diaconis (2013)] in which the generated graphs are almost full or empty. To overcome this disadvantage, Wang et al. (2009) proposes a number of new two-mode specifications such as the  $k$ -two-paths and three-path statistics. Although Wang et al. (2009) show that these model specifications have good performance by simulations, the theoretical properties of the model are still unknown.

## Appendix

In this appendix, we will present the proofs for Theorems 1 and 2. We start with some preliminaries. For an  $n \times n$  matrix  $J = (J_{i,j})$ ,  $\|J\|_\infty$  denotes the matrix norm induced by the  $\|\cdot\|_\infty$ -norm on vectors in  $\mathbb{R}^n$ :

$$\|J\|_\infty = \max_{\mathbf{x} \neq 0} \frac{\|J\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |J_{i,j}|.$$

Let  $D$  be an open convex subset of  $\mathbb{R}^n$ . We say an  $n \times n$  function matrix  $F(\mathbf{x})$  whose elements  $F_{ij}(\mathbf{x})$  are functions on vectors  $\mathbf{x}$ , is Lipschitz continuous on  $D$  if there exists a real number  $\lambda$  such that for any  $\mathbf{v} \in \mathbb{R}^n$  and any  $\mathbf{x}, \mathbf{y} \in D$ ,

$$\|F(\mathbf{x})(\mathbf{v}) - F(\mathbf{y})(\mathbf{v})\|_\infty \leq \lambda \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty,$$

where  $\lambda$  may depend on  $n$  but independent of  $\mathbf{x}$  and  $\mathbf{y}$ . For fixed  $n$ ,  $\lambda$  is a constant.

We introduce a class of matrices. Given two positive numbers  $q, Q$ , we say the  $(m+n-1) \times (m+n-1)$  matrix  $V = (v_{i,j})$  belongs to the class  $\mathcal{L}_{m,n}(q, Q)$  if the following holds:

$$\begin{aligned} q &\leq v_{i,i} - \sum_{j=m+1}^{m+n-1} v_{i,j} \leq Q, i = 1, \dots, m; v_{m,m} = \sum_{j=m+1}^{m+n-1} v_{m,j}, \\ v_{i,j} &= 0, i, j = 1, \dots, m, i \neq j, \\ v_{i,j} &= 0, i, j = m+1, \dots, m+n-1, i \neq j, \\ q &\leq v_{i,j} = v_{j,i} \leq Q, i = 1, \dots, m, j = m+1, \dots, m+n-1, \\ v_{i,i} &= \sum_{k=1}^m v_{k,i} = \sum_{k=1}^m v_{i,k}, i = m+1, \dots, m+n-1. \end{aligned} \tag{5.1}$$

If  $V \in \mathcal{L}_{m,n}(q, Q)$ , then  $V$  is a  $(m+n-1) \times (m+n-1)$  diagonally dominant, symmetric nonnegative matrix. Define  $v_{m+n,i} = v_{i,m+n} := v_{i,i} - \sum_{j=1}^{m+n-1} v_{i,j}$  for  $i = 1, \dots, m+n-1$  and  $v_{m+n,m+n} = \sum_{i=1}^{m+n-1} v_{m+n,i}$ . Then  $q \leq v_{m+n,i} \leq Q$  for  $i = 1, \dots, m$ ,  $v_{m+n,i} = 0$  for  $i = m, m+1, \dots, m+n-1$  and  $v_{m+n,m+n} = \sum_{i=1}^m v_{i,m+n} = \sum_{i=1}^m v_{m+n,i}$ . The Fisher information matrix of the parameter vector  $\boldsymbol{\theta}$ ,  $V$ , belongs to the matrix class  $V \in \mathcal{L}_{m,n}(q, Q)$ . The approximate error using  $S$  in (3.2) to approximate the inverse of  $V$  is given in the lemma



below, whose proof is the extension of that for Proposition 1 in [Yan et al. \(2016a\)](#).

**Lemma 1.** *If  $V \in \mathcal{L}_{m,n}(q, Q)$  with  $Q/q = o(n)$  and  $m/n = O(1)$ , then for large enough  $n$ ,*

$$\|V^{-1} - S\| \leq \frac{c_1 Q^2}{q^3 m n},$$

where  $c_1$  is a constant that dose not depend on  $q, Q, m$  and  $n$ , and  $\|A\| := \max_{i,j} |a_{i,j}|$  for a general matrix  $A = (a_{i,j})$ .

*Proof.* Recall that if  $V \in \mathcal{L}_{m,n}(q, Q)$ , then for  $i = 1, \dots, m+n$ ,

$$\begin{aligned} v_{i,i} &= \sum_{j=1}^{m+n} (1 - \delta_{i,j}) v_{i,j} = \sum_{j=1}^{m+n} (1 - \delta_{j,i}) v_{j,i} \\ &= \begin{cases} \sum_{j=m+1}^{m+n} v_{i,j} = \sum_{j=m+1}^{m+n} v_{j,i}, & i = 1, \dots, m, \\ \sum_{j=1}^m v_{i,j} = \sum_{j=1}^m v_{j,i}, & i = m+1, \dots, m+n, \end{cases} \end{aligned}$$

and if  $v_{i,j} > 0$  for  $1 \leq i \leq m, 1 \leq j \leq n, i \neq j$ , then

$$q \leq v_{i,j} \leq Q.$$

The above equation and inequality will be repeatedly used in this proof.

Let  $I$  denote the  $(m+n-1) \times (m+n-1)$  identity matrix. Define  $F = (f_{i,j}) = V^{-1} - S$ ,  $U = (u_{i,j}) = I - VS$  and  $W = (w_{i,j}) = SU$ . Then we have the recursion

$$F = T^{-1} - S = (T^{-1} - S)(I - TS) + S(I - TS) = FU + W. \quad (5.2)$$

Note that for  $i = 1, \dots, m, j = 1, \dots, m$ , we have

$$\begin{aligned} u_{i,j} &= \delta_{i,j} - \sum_{k=1}^{m+n-1} v_{i,k} s_{k,j} \\ &= \delta_{i,j} - \left[ \sum_{k=1}^m v_{i,k} \left( \frac{\delta_{k,j}}{v_{j,j}} + \frac{1}{v_{m+n,m+n}} \right) + \sum_{k=m+1}^{m+n-1} v_{i,k} \left( \frac{\delta_{k,j}}{v_{j,j}} - \frac{1}{v_{m+n,m+n}} \right) \right] \\ &= (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{j,j}} - \frac{v_{i,i} - \sum_{k=m+1}^{m+n-1} v_{i,k}}{v_{m+n,m+n}} = (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{j,j}} - \frac{v_{i,m+n}}{v_{m+n,m+n}}. \end{aligned}$$

Similarly, we also have

$$u_{i,j} = \begin{cases} (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{j,j}} + \frac{v_{i,m+n}}{v_{m+n,m+n}} & i = 1, \dots, m; j = m+1, \dots, m+n-1, \\ (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{j,j}} & i = m+1, \dots, m+n-1; j = 1, \dots, m, m+1, \dots, m+n-1. \end{cases}$$

In all,  $u_{i,j}$  can be written in a unified form:

$$u_{i,j} = (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{j,j}} + (1_{\{i \leq m, j > m\}} - 1_{\{i \leq m, j \leq m\}}) \frac{v_{i,m+n}}{v_{m+n,m+n}}, \quad (5.3)$$

where  $1_{\{\cdot\}}$  is an indicator function. Similarly, for  $i = 1, \dots, m$ ,  $j = 1, \dots, m$ , we have

$$\begin{aligned} w_{i,j} &= \sum_{k=1}^{m+n-1} s_{i,k} u_{k,j} \\ &= \sum_{k=1}^m \left( \frac{\delta_{i,k}}{v_{i,i}} + \frac{1}{v_{m+n,m+n}} \right) \left[ (\delta_{k,j} - 1) \frac{v_{k,j}}{v_{j,j}} - \frac{v_{k,m+n}}{v_{m+n,m+n}} \right] + \sum_{k=m+1}^{m+n-1} \left( -\frac{1}{v_{m+n,m+n}} \right) \left[ (\delta_{k,j} - 1) \frac{v_{k,j}}{v_{j,j}} \right] \\ &= \frac{(\delta_{i,j} - 1) v_{i,j}}{v_{i,i} v_{j,j}} - \frac{v_{i,m+n}}{v_{i,i} v_{m+n,m+n}} + 0 - \frac{v_{m+n,m+n}}{v_{m+n,m+n}^2} + \frac{v_{j,j} - v_{m+n,j}}{v_{m+n,m+n} v_{j,j}} \\ &= \frac{(\delta_{i,j} - 1) v_{i,j}}{v_{i,i} v_{j,j}} - \frac{v_{i,m+n}}{v_{i,i} v_{m+n,m+n}} - \frac{v_{m+n,j}}{v_{m+n,m+n} v_{j,j}}, \end{aligned}$$

and

$$w_{i,j} = \begin{cases} (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{i,i} v_{j,j}} + \frac{v_{i,m+n}}{v_{i,i} v_{m+n,m+n}} & i = 1, \dots, m; j = m+1, \dots, m+n-1, \\ (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{i,i} v_{j,j}} + \frac{v_{m+n,j}}{v_{j,j} v_{m+n,m+n}} & i = m+1, \dots, m+n-1; j = 1, \dots, m, \\ (\delta_{i,j} - 1) \frac{v_{i,j}}{v_{i,i} v_{j,j}} & i = m+1, \dots, m+n-1; j = m+1, \dots, m+n-1. \end{cases}$$

Further, when  $1 \leq i \neq j \leq m+n$ ,

$$0 \leq \frac{v_{i,j}}{v_{i,i} v_{j,j}} \leq \frac{Q}{q^2 m n},$$

and it is not difficult to show that, when  $i, j, k$  are different from each other, we have

$$\begin{aligned} |w_{i,i}| &\leq \frac{2Q}{q^2 m n}, \\ |w_{i,j}| &\leq \frac{3Q}{q^2 m n}, \\ |w_{i,j} - w_{i,k}| &\leq \frac{2Q}{q^2 m n}, \\ |w_{i,i} - w_{i,k}| &\leq \frac{2Q}{q^2 m n}. \end{aligned}$$

It follows that

$$\max(|w_{i,j}|, |w_{i,j} - w_{i,k}|) \leq \frac{3Q}{q^2 m n} \quad \text{for all } i, j, k. \quad (5.4)$$

Next we use the recursion (5.2) to obtain an upper bound of the approximate error  $\|F\|$ .

By (5.2) and (5.3), for any  $i \in \{1, \dots, m+n-1\}$ , we have that for  $j = 1, \dots, m+n-1$

$$f_{i,j} = \sum_{k=1}^{m+n-1} f_{i,k} [(\delta_{k,j} - 1) \frac{v_{k,j}}{v_{j,j}} + (1_{\{k \leq m, j > m\}} - 1_{\{k \leq m, j \leq m\}}) \frac{v_{k,m+n}}{v_{m+n,m+n}}] + w_{i,j}.$$

Thus, to prove Lemma 1, it is sufficient to show that  $|f_{i,j}| \leq c_1 Q^2 / (q^3 mn)$  for any  $i, j$ . The condition  $m/n = O(1)$  guarantees that  $Q/q = o(m)$  if  $Q/q = o(n)$ . The remainder of the proof of Lemma 1 is similar to the proof of Proposition 1 in Yan et al. (2016a), and we omit the details here.  $\square$

Note that if  $Q$  and  $q$  are bounded constants, then the upper bound of the above approximation error is on the order of  $(mn)^{-1}$ , indicating that  $S$  is a high-accuracy approximation to  $V^{-1}$ . Further, based on the above proposition, we immediately have the following lemma.

**Lemma 2.** *If  $V \in \mathcal{L}_n(q, Q)$  with  $Q/q = o(n)$  and  $m/n = O(1)$ , then for a vector  $\mathbf{x} \in \mathbb{R}^{m+n-1}$ ,*

$$\|V^{-1}\mathbf{x}\|_{\infty} \leq \frac{2c_1 Q^2}{q^3 mn} + \frac{|x_{m+n}|}{v_{m+n,m+n}} + \max_{i=1, \dots, m+n-1} \frac{|x_i|}{v_{i,i}},$$

where  $x_{m+n} := \sum_{i=1}^m x_i - \sum_{i=m+1}^{m+n-1} x_i$ .

Similar to Theorem 8 in Yan et al. (2016a), we have the following lemma for the rate of convergence for the Newton's iterative sequence to solve the system of the likelihood equations, whose proof is similar to that in Yan et al. (2016a) and we omit it here.

**Lemma 3.** *Define a system of equations:*

$$\begin{aligned} F_i(\boldsymbol{\theta}) &= d_i - \sum_{k=1}^n f(\alpha_i + \beta_k), i = 1, \dots, m, \\ F_{m+j}(\boldsymbol{\theta}) &= b_j - \sum_{k=1}^m f(\alpha_k + \beta_j), j = 1, \dots, n-1, \\ F(\boldsymbol{\theta}) &= (F_1(\boldsymbol{\theta}), \dots, F_m(\boldsymbol{\theta}), F_{m+1}(\boldsymbol{\theta}), \dots, F_{m+n-1}(\boldsymbol{\theta}))^T, \end{aligned}$$

where  $f(\cdot)$  is a continuous function with the third derivative. Let  $D \subset \mathbb{R}^{m+n-1}$  be a convex set and assume for any  $\mathbf{x}, \mathbf{y}, \mathbf{v} \in D$ , we have

$$\| [F'(\mathbf{x}) - F'(\mathbf{y})] \mathbf{v} \|_{\infty} \leq K_1 \| \mathbf{x} - \mathbf{y} \|_{\infty} \| \mathbf{v} \|_{\infty},$$

$$\max_{i=1, \dots, m+n-1} \| F'_i(\mathbf{x}) - F'_i(\mathbf{y}) \|_{\infty} \leq K_2 \| \mathbf{x} - \mathbf{y} \|_{\infty},$$

where  $F'(\boldsymbol{\theta})$  is the Jacobian matrix of  $F$  on  $\boldsymbol{\theta}$  and  $F'_i(\boldsymbol{\theta})$  is the gradient function of  $F_i$  on  $\boldsymbol{\theta}$ . Consider  $\boldsymbol{\theta}^{(0)} \in D$  with  $\Omega(\boldsymbol{\theta}^{(0)}, 2r) \subset D$  where  $r = \| [F'(\boldsymbol{\theta}^{(0)})]^{-1} F(\boldsymbol{\theta}^{(0)}) \|_{\infty}$  for any

$\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$ . We assume that  $m/n = O(1)$  and

$$F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(q, Q) \quad \text{or} \quad -F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(q, Q).$$

For  $k = 1, 2, \dots$ , define the Newton iterates  $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [F'(\boldsymbol{\theta}^{(k)})]^{-1}F(\boldsymbol{\theta}^{(k)})$ . Let

$$\rho = \frac{c_1(m+n-1)Q^2K_1}{2q^3mn} + \frac{K_2}{mq}.$$

If  $\rho < 1/2$ , then  $\boldsymbol{\theta}^{(k)} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$ ,  $k = 1, 2, \dots$ , are well defined and satisfy

$$\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(0)}\|_{\infty} \leq r/(1 - \rho r).$$

Further,  $\lim_{k \rightarrow \infty} \boldsymbol{\theta}^{(k)}$  exists and the limiting point is precisely the solution of  $F(\boldsymbol{\theta}) = 0$  in the range of  $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$ .

## Appendix A: Proofs for Theorem 1

We define a system of functions:

$$\begin{aligned} F_i(\boldsymbol{\theta}) &= d_i - \sum_{j=1}^n \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad i = 1, \dots, m, \\ F_{m+j}(\boldsymbol{\theta}) &= b_j - \sum_{i=1}^m \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}, \quad j = 1, \dots, n-1 \\ F(\boldsymbol{\theta}) &= (F_1(\boldsymbol{\theta}), \dots, F_{m+n-1}(\boldsymbol{\theta}))^\top. \end{aligned}$$

Note that the solution to the equation  $F(\boldsymbol{\theta}) = 0$  is precisely the MLE. Then the Jacobin matrix  $F'(\boldsymbol{\theta})$  of  $F(\boldsymbol{\theta})$  can be calculated as follows. For  $i = 1, \dots, m$ ,

$$\frac{\partial F_i}{\partial \alpha_l} = 0, l = 1, \dots, m, l \neq i; \quad \frac{\partial F_i}{\partial \alpha_i} = - \sum_{j=1}^n \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2},$$

$$\frac{\partial F_i}{\partial \beta_j} = - \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}, \quad j = 1, \dots, n-1,$$

and for  $j = 1, \dots, n-1$ ,

$$\frac{\partial F_{m+j}}{\partial \alpha_l} = - \frac{e^{\alpha_l + \beta_j}}{(1 + e^{\alpha_l + \beta_j})^2}, \quad l = 1, \dots, m,$$

$$\frac{\partial F_{m+j}}{\partial \beta_j} = - \sum_{i=1}^m \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}; \quad \frac{\partial F_{m+j}}{\partial \beta_k} = 0, \quad k = 1, \dots, n-1, k \neq j.$$

Since  $e^x/(1+e^x)^2$  is a decreasing function on  $x$  when  $x \geq 0$  and an increasing function when  $x \leq 0$ . Consequently, for any  $i, j$ , we have

$$\frac{e^{2\|\boldsymbol{\theta}\|_\infty}}{(1+e^{2\|\boldsymbol{\theta}\|_\infty})^2} \leq -F'_{i,j}(\boldsymbol{\theta}) \leq \frac{1}{4}.$$

According to the definition of  $\mathcal{L}_{m,n}(q, Q)$ , we have that  $-F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(q, Q)$ , where

$$q = \frac{e^{2\|\boldsymbol{\theta}\|_\infty}}{(1+e^{2\|\boldsymbol{\theta}\|_\infty})^2}, \quad Q = \frac{1}{4}.$$

Therefore, Lemma 3 and Lemma 1 can be applied. Let  $\boldsymbol{\theta}^*$  denote the true parameter vector. The constants  $K_1, K_2$  and  $r$  in the upper bounds of Lemma 3 are given in the following lemma.

**Lemma 4.** Take  $D = R^{m+n-1}$  and  $\boldsymbol{\theta}^0 = \boldsymbol{\theta}^*$  in Lemma 3. Assume

$$\max\left\{\max_{i=1,\dots,m} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\dots,n} |b_j - \mathbb{E}(b_j)|\right\} \leq \sqrt{n \log n}. \quad (\text{A1})$$

If  $m/n = O(1)$ , then we can choose the constants  $K_1, K_2$  and  $r$  in Lemma 3 as

$$K_1 = n, K_2 = \frac{n}{2}, r \leq \frac{(\log n)^{1/2}}{n^{1/2}} (c_{11}e^{6\|\boldsymbol{\theta}^*\|_\infty} + c_{12}e^{2\|\boldsymbol{\theta}^*\|_\infty}),$$

where  $c_{11}, c_{12}$  are constants.

*Proof.* For fixed  $m, n$ , we first derive  $K_1$  and  $K_2$  in the inequalities of Lemma 3. Let  $\mathbf{x}, \mathbf{y} \in R^{m+n-1}$  and

$$F'_i(\boldsymbol{\theta}) = (F'_{i,1}(\boldsymbol{\theta}), \dots, F'_{i,m+n-1}(\boldsymbol{\theta})) := \left(\frac{\partial F_i}{\partial \alpha_1}, \dots, \frac{\partial F_i}{\partial \alpha_m}, \frac{\partial F_i}{\partial \beta_1}, \dots, \frac{\partial F_i}{\partial \beta_{n-1}}\right).$$

Then, for  $i = 1, \dots, m$ , we have

$$\begin{aligned} \frac{\partial^2 F_i}{\partial \alpha_l \partial \alpha_s} &= 0, s \neq l; -\sum_{j=1}^n \frac{e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3}, \\ \frac{\partial^2 F_i}{\partial \alpha_i \partial \beta_s} &= -\frac{e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3}, s = 1, \dots, n-1, s \neq i; \frac{\partial^2 F_i}{\partial \alpha_i \partial \beta_i} = 0, \\ \frac{\partial^2 F_i}{\partial \beta_j^2} &= -\frac{e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3}, j = 1, \dots, n-1; \frac{\partial^2 F_i}{\partial \beta_s \partial \beta_l} = 0, s \neq l. \end{aligned}$$

Note that

$$\left| \frac{e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3} \right| \leq \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2} \leq \frac{1}{4}. \quad (\text{A2})$$

By the mean value theorem for vector-valued functions (Lang (1993), p.341), we have

$$F'_i(\mathbf{x}) - F'_i(\mathbf{y}) = J^{(i)}(\mathbf{x} - \mathbf{y}),$$

where

$$J_{s,l}^{(i)} = \int_0^1 \frac{\partial F'_{i,s}}{\partial \boldsymbol{\theta}_l}(t\mathbf{x} + (1-t)\mathbf{y})dt, \quad s, l = 1, \dots, m+n-1.$$

Therefore,

$$\max_s \sum_l^{m+n-1} |J^{(i)}| \leq \frac{n}{2}, \quad \sum_{s,l} |J_{(s,l)}^{(i)}| \leq n$$

Similarly, for  $i = m+1, \dots, m+n-1$ , we also have  $F'_i(\mathbf{x}) - F'_i(\mathbf{y}) = J^{(i)}(\mathbf{x} - \mathbf{y})$  and  $\sum_{s,l} |J_{(s,l)}^{(i)}| \leq m$ . Consequently,

$$\|F'_i(\mathbf{x}) - F'_i(\mathbf{y})\|_\infty \leq \|J^{(i)}\|_\infty \|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{n}{2} \|\mathbf{x} - \mathbf{y}\|_\infty, \quad i = 1, \dots, m+n-1,$$

and for  $\forall \mathbf{v} \in R^{m+n-1}$ ,

$$\begin{aligned} \|[F'_i(\mathbf{x}) - F'_i(\mathbf{y})]\mathbf{v}\|_\infty &= \max_i \left| \sum_{j=1}^{m+n-1} (F'_{i,j}(\mathbf{x}) - F'_{i,j}(\mathbf{y}))v_j \right| \\ &= \max_i |(\mathbf{x} - \mathbf{y})J^{(i)}\mathbf{v}| \\ &\leq \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty \sum_{k,j} |J_{(s,l)}^{(i)}| \\ &\leq n \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty. \end{aligned}$$

So we can choose  $K_1 = n$  and  $K_2 = n/2$  in Lemma 3.

It is obvious that  $-F'(\boldsymbol{\theta}^*) \in \mathcal{L}_n(q_*, Q_*)$ , where

$$q_* = \frac{e^{2\|\boldsymbol{\theta}^*\|_\infty}}{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2}, \quad Q_* = \frac{1}{4}.$$

Note that

$$F(\boldsymbol{\theta}^*) = (d_1 - \mathbb{E}(d_1), \dots, d_m - \mathbb{E}(d_m), b_1 - \mathbb{E}(b_1), \dots, b_{n-1} - \mathbb{E}(b_{n-1})).$$

By the assumption of A1 and Lemma 2, if  $m/n = O(1)$ , then we have

$$\begin{aligned} r = \|[F'(\boldsymbol{\theta}^*)]^{-1}F(\boldsymbol{\theta}^*)\|_\infty &\leq \frac{2c_1(m+n-1)Q_*^2\|F(\boldsymbol{\theta}^*)\|_\infty}{q_*^2mn} + \max_{i=1,\dots,m+n-1} \frac{|F_i(\boldsymbol{\theta}^*)|}{v_{i,i}} + \frac{|F_{m+n}(\boldsymbol{\theta}^*)|}{v_{m+n,m+n}} \\ &\leq \frac{(\log n)^{1/2}}{n^{1/2}} (c_{11}e^{6\|\boldsymbol{\theta}^*\|_\infty} + c_{12}e^{2\|\boldsymbol{\theta}^*\|_\infty}), \end{aligned}$$

where  $c_{11}, c_{12}$  are constants. □

The following lemma assures that condition (A1) holds with a large probability.

**Lemma 5.** *With probability at least  $1 - 4/n$ , we have*

$$\max\{\max_{i=1,\dots,m} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\dots,n} |b_j - \mathbb{E}(b_j)|\} \leq \sqrt{n \log n}.$$

*Proof.* Note that  $m < n$ . By Hoeffding (1963)'s inequality, we have

$$P\left(|d_i - \mathbb{E}(d_i)| \geq \sqrt{n \log n}\right) \leq 2 \exp\left\{-\frac{2n \log n}{m}\right\} \leq \frac{2}{n^{2n/m}} \leq \frac{2}{n^2}.$$

Therefore,

$$\begin{aligned} P\left(\max_i |d_i - \mathbb{E}(d_i)| \geq \sqrt{n \log n}\right) &\leq P\left(\bigcup_i |d_i - \mathbb{E}(d_i)| \geq \sqrt{n \log n}\right) \\ &\leq \sum_{i=1}^m P\left(|d_i - \mathbb{E}(d_i)| \geq \sqrt{n \log n}\right) \\ &\leq m \times \frac{2}{n^2} \leq \frac{2}{n}. \end{aligned}$$

Similarly, we have

$$P\left(\max_j |b_j - \mathbb{E}(b_j)| \geq \sqrt{n \log n}\right) \leq \frac{2}{n}.$$

Consequently,

$$\begin{aligned} &P\left(\max\{\max_{i=1,\dots,m} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\dots,n} |b_j - \mathbb{E}(b_j)|\} \geq \sqrt{n \log n}\right) \\ &\leq P\left(\max_i |d_i - \mathbb{E}(d_i)| \geq \sqrt{n \log n}\right) + P\left(\max_j |b_j - \mathbb{E}(b_j)| \geq \sqrt{n \log n}\right) \\ &\leq \frac{4}{n}. \end{aligned}$$

This is equivalent to Lemma 5. □

**Proof of Theorem 1.** Assume that condition (A1) holds. Recall that the Newton's iterates in Lemma 3,  $\boldsymbol{\theta}^{(k+1)} = [F'(\boldsymbol{\theta}^{(k)})]^{-1} F(\boldsymbol{\theta}^{(k)})$  with  $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^*$ . If  $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$ , then  $-F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}(q, Q)$  with

$$q = \frac{e^{2\|\boldsymbol{\theta}^*\|_\infty + 2r}}{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty + 2r})^2}, \quad Q = \frac{1}{4}.$$

By Lemma 4 and condition (A1), for sufficient small  $r$ ,

$$\begin{aligned} \rho r &\leq \left[ \frac{c_1(m+n-1)Q^2n}{2q^3mn} + \frac{n}{2nq} \right] \times \frac{(\log n)^{1/2}}{n^{1/2}} (c_{11}e^{6\|\boldsymbol{\theta}^*\|_\infty} + c_{12}e^{2\|\boldsymbol{\theta}^*\|_\infty}) \\ &\leq O\left(\frac{(\log n)^{1/2}e^{12\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}}\right) + O\left(\frac{(\log n)^{1/2}e^{8\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}}\right). \end{aligned}$$

If  $\boldsymbol{\theta}^* \in \mathbb{R}^{m+n-1}$  with  $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ , where  $0 < \tau < 1/24$  is a constant, then as  $n \rightarrow \infty$ ,

$$(\log n)^{1/2} n^{-1/2} e^{12\|\boldsymbol{\theta}^*\|_\infty} \leq (\log n)^{1/2} n^{-1/2+12\tau} \rightarrow 0.$$

Therefore,  $\rho r \rightarrow 0$  as  $n \rightarrow \infty$ . By Lemma 3,  $\lim_{n \rightarrow \infty} \hat{\boldsymbol{\theta}}^{(n)}$  exists. Denote the limit as  $\hat{\boldsymbol{\theta}}$ . Then it satisfies

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \leq 2r = O\left(\frac{(\log n)^{1/2} e^{6\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}}\right) = o(1).$$

By Lemma 5, condition (A1) holds with probability approaching to one, thus the above inequality also holds with probability approaching to one. The uniqueness of the MLE comes from Proposition 5 in Yan et al. (2016a).  $\square$

## Appendix B: Proofs for Theorem 2

We first present one proposition. Since  $d_i = \sum_k a_{i,k}$  and  $b_j = \sum_k a_{k,j}$  are sums of  $m$  and  $n$  independent random variables, by the central limit theorem for the bounded case in Loève (Loeve (1977), page 289), we know that  $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$  and  $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$  are asymptotically standard normal if  $v_{i,i}$  and  $v_{m+j,m+j}$  diverge, respectively. Note that

$$\frac{me^{2\|\boldsymbol{\theta}^*\|_\infty}}{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2} \leq v_{i,i} \leq \frac{m}{4}, \quad \frac{ne^{2\|\boldsymbol{\theta}^*\|_\infty}}{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2} \leq v_{m+j,m+j} \leq \frac{n}{4}.$$

Then we have the following proposition.

**Proposition 1.** *Assume that  $X \sim \mathbb{P}_{\boldsymbol{\theta}^*}$ . If  $e^{\|\boldsymbol{\theta}^*\|_\infty} = o(n^{1/2})$ , then for any fixed  $k \geq 1$ , as  $n \rightarrow \infty$ , the vector consisting of the first  $k$  elements of  $S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$  is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left  $k \times k$  block of  $S$ .*

To complete the proof of Theorem 2, we need two lemmas as follows.

**Lemma 6.** *Let  $R = V^{-1} - S$  and  $U = \text{Cov}[R\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}]$ . Then*

$$\|U\| \leq \|V^{-1} - S\| + \frac{3(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^4}{4mne^{4\|\boldsymbol{\theta}^*\|_\infty}}.$$

*Proof.* Note that

$$U = RVR^T = (V^{-1} - S)V(V^{-1} - S)^T = (V^{-1} - S) - S(I - VS),$$

where  $I$  is a  $(m + n - 1) \times (m + n - 1)$  diagonal matrix, and by (5.4), we have

$$|\{S(I - VS)\}_{i,j}| = |w_{i,j}| \leq \frac{3(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^4}{4mne^{4\|\boldsymbol{\theta}^*\|_\infty}}.$$



Thus,

$$\|U\| \leq \|V^{-1} - S\| + \|\{S(I_{m+n-1} - VS)\}\| \leq \|V^{-1} - S\| + \frac{3(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^4}{4mne^{4\|\boldsymbol{\theta}^*\|_\infty}}.$$

□

**Lemma 7.** Assume that the conditions in Theorem 1 hold. If  $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$  with  $\tau < 1/24$  and  $m/n = O(1)$ , then for any  $i$ ,

$$\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i^* = [V^{-1}\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i + o_p(n^{-1/2}).$$

*Proof.* By Theorem 1, if  $m/n = O(1)$ , then we have

$$\widehat{\rho}_n := \max_{1 \leq i \leq m+n-1} |\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i^*| = O_p\left(\frac{(\log n)^{1/2} e^{6\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}}\right).$$

Let  $\widehat{\gamma}_{i,j} = \widehat{\alpha}_i + \widehat{\beta}_j - \alpha_i^* - \beta_j^*$ . By the Taylor expansion, for any  $1 \leq i \leq m, 1 \leq j \leq n, i \neq j$ ,

$$\frac{e^{\widehat{\alpha}_i + \widehat{\beta}_j}}{1 + e^{\widehat{\alpha}_i + \widehat{\beta}_j}} - \frac{e^{\alpha_i^* + \beta_j^*}}{1 + e^{\alpha_i^* + \beta_j^*}} = \frac{e^{\alpha_i^* + \beta_j^*}}{(1 + e^{\alpha_i^* + \beta_j^*})^2} \widehat{\gamma}_{i,j} + h_{i,j},$$

where

$$h_{i,j} = \frac{e^{\alpha_i^* + \beta_j^* + \phi_{i,j} \widehat{\gamma}_{i,j}} (1 - e^{\alpha_i^* + \beta_j^* + \phi_{i,j} \widehat{\gamma}_{i,j}})}{2(e^{\alpha_i^* + \beta_j^* + \phi_{i,j} \widehat{\gamma}_{i,j}})^3} \widehat{\gamma}_{i,j}^2,$$

and  $0 \leq \phi_{i,j} \leq 1$ . By the likelihood equations (3.1), it is not difficult to verify that

$$\mathbf{g} - \mathbb{E}\mathbf{g} = V(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + \mathbf{h},$$

where  $\mathbf{h} = (h_1, \dots, h_{m+n-1})^T$  and

$$h_i = \sum_{k=1}^n h_{i,k}, i = 1, \dots, m, \quad h_{m+i} = \sum_{k=1}^m h_{k,i}, i = 1, \dots, n-1.$$

Equivalently,

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = V^{-1}(\mathbf{g} - \mathbb{E}\mathbf{g}) + V^{-1}\mathbf{h}. \quad (\text{C1})$$

By (A2), it is easy to show

$$|h_{i,j}| \leq |\widehat{\gamma}_{i,j}^2/2| \leq 2\widehat{\rho}_n^2, \quad |h_i| \leq \sum_{i,j} |h_{i,j}| \leq 2n\widehat{\rho}_n^2.$$

Because

$$h_{m+n} = \sum_{i=1}^m h_i - \sum_{j=1}^{n-1} h_{m+j} = \sum_{i=1}^m \sum_{j=1}^n h_{i,j} - \sum_{j=1}^{n-1} \sum_{i=1}^m h_{i,j} = \sum_{i=1}^m h_{i,n}.$$

Therefore

$$|h_{m+n}| \leq (m+n)\widehat{\rho}_n^2.$$

Note that  $(S\mathbf{h})_i = h_i/v_{i,i} + (-1)^{1_{\{i>m\}}} h_{m+n}/v_{m+n,m+n}$ , and  $(V^{-1}\mathbf{h})_i = (S\mathbf{h})_i + (R\mathbf{h})_i$ . Then we have

$$|(S\mathbf{h})_i| \leq \frac{|h_i|}{v_{i,i}} + \frac{|h_{2n}|}{v_{m+n,m+n}} \leq \frac{4\widehat{\rho}_n^2 \cdot (1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2}{e^{2\|\boldsymbol{\theta}^*\|_\infty}} \leq O\left(\frac{e^{14\|\boldsymbol{\theta}^*\|_\infty} \log n}{n}\right),$$

by Lemma 1, we have

$$|(R\mathbf{h})_i| \leq \|R\|_\infty \times [(m+n-1) \max_i |h_i|] \leq O\left(\frac{e^{18\|\boldsymbol{\theta}^*\|_\infty} \log n}{n}\right).$$

If  $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ , and  $\tau < 1/36$ , then  $|(V^{-1}\mathbf{h})_i| \leq |(S\mathbf{h})_i| + |(R\mathbf{h})_i| = o(n^{-1/2})$ . □

**Proof of Theorem 2.** By (C1), we have

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)_i = [S\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i + [R\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i + (V^{-1}\mathbf{h})_i.$$

By Lemma 6, we have

$$\|U\| \leq \|V^{-1} - S\| + \frac{3(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^4}{4mne^{4\|\boldsymbol{\theta}^*\|_\infty}} \leq O\left(\frac{e^{6\|\boldsymbol{\theta}^*\|_\infty}}{mn}\right) + O\left(\frac{e^{4\|\boldsymbol{\theta}^*\|_\infty}}{mn}\right) = O\left(\frac{e^{6\|\boldsymbol{\theta}^*\|_\infty}}{mn}\right).$$

By Chebyshev's inequality, if  $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ ,  $m/n = O(1)$ , and  $\tau < 1/36$ , then

$$P\left\{\frac{[R\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i}{n^{-1/2}} > \epsilon\right\} \leq \frac{\text{Cov}[R\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i}{n\epsilon^2} \leq \frac{1}{n\epsilon^2} O\left(\frac{e^{6\|\boldsymbol{\theta}^*\|_\infty}}{mn}\right) = o(1),$$

Therefore, we have

$$[R\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i = o_p(n^{-1/2}).$$

By Lemma 6 and Lemma 7, we have

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)_i = [S\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]_i + o_p(n^{-1/2}).$$

Theorem 2 follows directly from Proposition 1. □

## Acknowledgements

We are very grateful to one anonymous referee and the Editor for their valuable comments that have greatly improved the manuscript. Qin’s research is partially supported by the National Natural Science Foundation of China (No.11271147, 11471135). Yan’s research is partially supported by the National Natural Science Foundation of China (No.11401239) and the self-determined research funds of CCNU from the colleges’s basic research and operation of MOE (CCNU15A02032, CCNU15ZD011) and a fund from KLAS (130026507). Zhang’s research is supported by a fund from CCNU (2016CXZZ157).

## References

- Bhattacharyya, S. and Bickel, P. J. (2016). Spectral clustering and block models: A review and a new algorithm. *Statistical Analysis for High-Dimensional Data*, pages 67–90.
- Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461.
- Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.*, 21(4):1400–1435.
- Canyon, M. J. and Muldoon, M. R. (2004). The small world network structure of boards of directors. Available at SSRN 546963, <http://ssrn.com/abstract=546963>.
- Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, 5:815–841.
- Hillar, C. and Wibisono, A. (2013). Maximum entropy distributions on graphs. Available at: <http://arxiv.org/abs/1301.3321>.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.
- Holl, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373):33–65.
- Iacobucci, D. and Wasserman, S. (1990). Social networks with two sets of actors. *Psychometrika*, 55(4):707–720.
- Lang, S. (1993). Real and functional analysis. *New York:Springer-Verlag*.

- Latapy, M., Magnien, C., and Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social networks*, 30(1):31–48.
- Loeve, M. (1977). Probability theory. 4th ed. *New York:Springer-Verlag*.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. *Copenhagen: Paedagogiske Institut*.
- Robins, G. and Alexander, M. (2004). Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Computational & Mathematical Organization Theory*, 10(1):69–94.
- Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics*, 41(2):508–535.
- Snijders, T. A., Lomi, A., and Torló, V. J. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social networks*, 35(2):265–276.
- Wang, P., Sharpe, K., Robins, G. L., and Pattison, P. E. (2009). Exponential random graph ( $p^*$ ) models for affiliation networks. *Social Networks*, 31(1):12–25.
- Yan, T., Leng, C., Zhu, J., et al. (2016a). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, 44(1):31–57.
- Yan, T., Qin, H., and Wang, H. (2016b). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statistica Sinica*, (26):273–293.
- Yan, T. and Xu, J. (2013). A central limit theorem in the  $\beta$ -model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2):519–524.

Figure 1: The QQ plots of  $\hat{v}_{ii}^{-1/2}(\hat{\alpha}_i - \alpha_i)$  ( $m = 100, n = 200$ ).

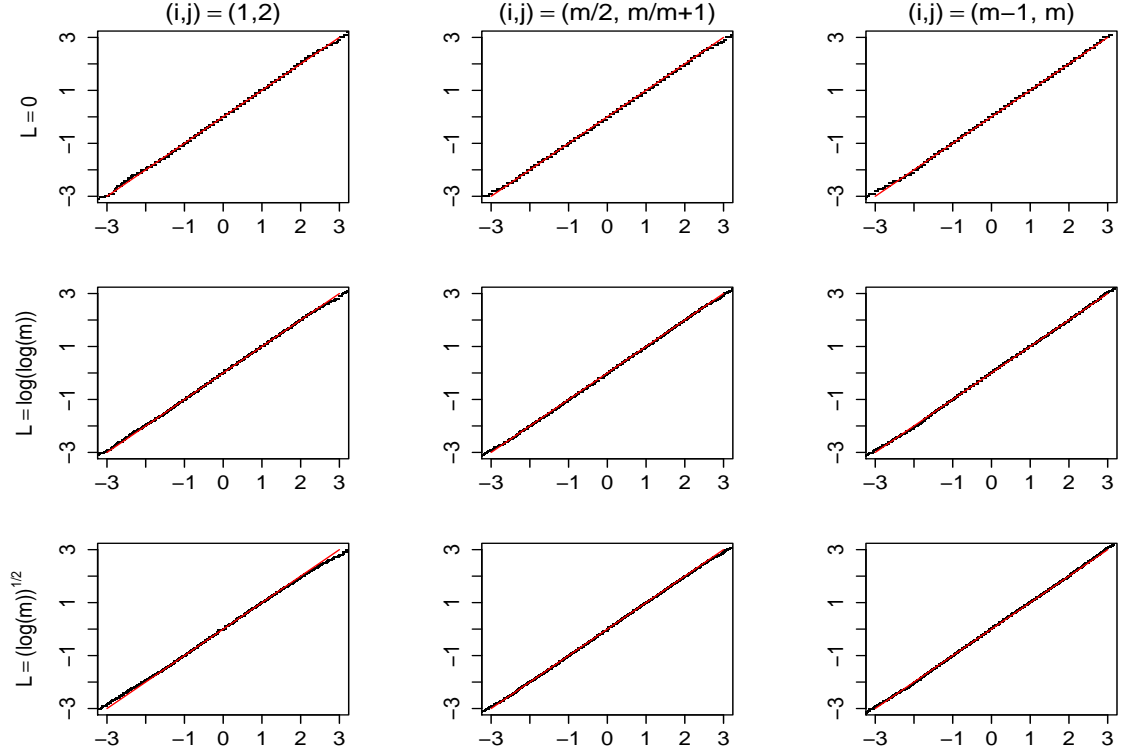


Figure 2: The QQ plots of  $\hat{v}_{jj}^{-1/2}(\hat{\beta}_j - \beta_j)$  ( $m = 100, n = 200$ ).

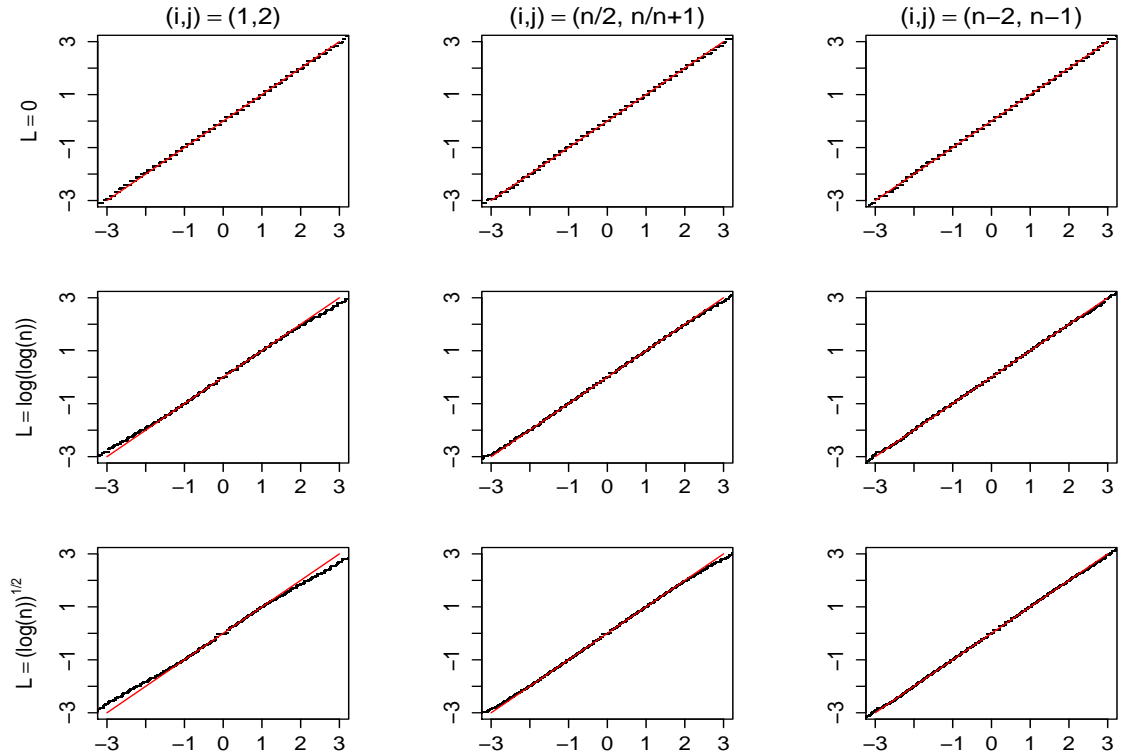


Table 1: Estimated coverage probabilities of  $\alpha_i - \alpha_j$  for pair  $(i, j)$  as well as the length of confidence intervals (in square brackets), and the probabilities that the MLE does not exist (in parentheses), multiplied by 100.

m	$(i, j)$	$L = 0$	$L = \log m$	$L = (\log m)^{1/2}$	$L = \log(m)$
100	(1, 2)	94.93[0.40](0)	95.17[0.55](0)	94.68[0.70](0.43)	(100)
	(50, 51)	95.15[0.40](0)	95.02[0.46](0)	94.78[0.52](0.43)	(100)
	(99, 100)	95.66[0.40](0)	95.17[0.42](0)	94.75[0.44](0.43)	(100)

Table 2: Estimated coverage probabilities of  $\beta_i - \beta_j$  for pair  $(i, j)$  as well as the length of confidence intervals (in square brackets), and the probabilities that the MLE does not exist (in parentheses), multiplied by 100.

n	$(i, j)$	$L = 0$	$L = \log n$	$L = (\log n)^{1/2}$	$L = \log(n)$
200	(1, 2)	94.77[0.60](0)	94.13[0.83](0)	94.75[1.17](2.41)	(100)
	(100, 101)	95.41[0.60](0)	94.79[0.68](0)	94.87[0.77](2.41)	(100)
	(198, 199)	96.60[0.60](0)	95.04[0.60](0)	94.86[0.63](2.41)	(100)

Table 3: The Student Extracurricular network data: the estimated influence parameters  $\hat{\theta}$ , 95% confidence intervals (in square brackets) and their standard errors (in parentheses)

Student Organizations	Degree	$\hat{\alpha}_i$
Spanish.Club	199	-0.32[-0.49, -0.16](0.08)
Pep.Club	157	-0.64[-0.82, -0.46](0.09)
NHS	124	-0.94[-1.14, -0.75](0.10)
Latin.Club	93	-1.29[-1.51, -1.07](0.11)
Orchestra..Symphonic	89	-1.34[-1.57, -1.12](0.11)
Key.Club	76	-1.52[-1.76, -1.28](0.12)
Spanish.Club..high.	68	-1.65[-1.90, -1.40](0.13)
Drunk.Driving	67	-1.67[-1.92, -1.41](0.13)
Forensics..National.Forensics.League.	66	-1.68[-1.94, -1.43](0.13)
Choir..a.capella	65	-1.70[-1.96, -1.44](0.13)
:	:	:
Chess.Club	7	-4.04[-4.78, -3.29](0.38)
Volleyball..JV	7	-4.04[-4.78, -3.29](0.38)
Teachers.of.Tomorrow	5	-4.38[-5.26, -3.50](0.45)
Quiz.Bowl..all.	5	-4.38[-5.26, -3.50](0.45)
Cheerleaders..Spirit.Squad	5	-4.38[-5.26, -3.50](0.45)
Drunk.Driving.Officers	4	-4.60[-5.58, -3.62](0.50)
Choir..vocal.ensemble..4.women.	4	-4.60[-5.58, -3.62](0.50)
Choir..barbershop.quartet..4.men.	4	-4.60[-5.58, -3.62](0.50)
Cross.Country..girls.8th	3	-4.89[-6.02, -3.76](0.58)
Swim...Dive.Team..boys	3	-4.89[-6.02, -3.76](0.58)

Table 4: The Student Extracurricular network data: the estimated influence parameters  $\hat{\theta}$ , 95% confidence intervals (in square brackets) and their standard errors (in parentheses)

Student ID	Degree	$\hat{\beta}_j$
122662	14	0.996[0.387, 1.606](0.311)
114850	12	0.792[0.146, 1.438](0.330)
888947	12	0.792[0.146, 1.438](0.330)
126259	11	0.680[0.011, 1.348](0.341)
139161	11	0.680[0.011, 1.348](0.341)
122638	10	0.559[−0.136, 1.254](0.355)
888981	10	0.559[−0.136, 1.254](0.355)
888988	10	0.559[−0.136, 1.254](0.355)
889059	10	0.559[−0.136, 1.254](0.355)
114037	9	0.428[−0.298, 1.153](0.370)
114671	9	0.428[−0.298, 1.153](0.370)
888892	9	0.428[−0.298, 1.153](0.370)
⋮	⋮	⋮
889102	1	−1.985[−3.974, 0.004](1.015)
889103	1	0.000[−0.844, 0.844](0.431)

The full table is available by sending emails to [zhang\\_yong@mails.ccn.edu.cn](mailto:zhang_yong@mails.ccn.edu.cn)